How Geoscience Novices Reason About Temporal Duration: The Role of Spatial Thinking and Large Numbers

Kim A. Cheek[1,a]

ABSTRACT

Research about geologic time conceptions generally focuses on the placement of events on the geologic timescale, with few studies dealing with the duration of geologic processes or events. Those studies indicate that students often have very poor conceptions about temporal durations of geologic processes, but the reasons for that are relatively unexplored. Close connections between ideas about time and space over short time periods, as well as poor number sense for numbers in unfamiliar ranges have been repeatedly demonstrated. This study explored whether the conceptions geoscience novices hold about the temporal duration of geoscience processes across a variety of temporal scales are influenced by their ideas about space and large numbers. Seventeen undergraduates in an introductory geoscience course participated in task-based qualitative interviews. Students tended to equate spatial size with temporal duration over short and long time periods, sometimes modifying contradictory data to fit their interpretation. Confusion about the relative size of temporal periods up to 100,000,000 years was observed. They described durations operating on long temporal scales in largely qualitative, imprecise terms. Spatial compression of large temporal periods and expansion of short time periods were common. Students possessed few specific temporal markers for durations of task events. Specific pedagogical and curricular recommendations are discussed. © *2013 National Association of Geoscience Teachers*. [DOI: 10.5408/12-365.1]

Key words: duration, geologic time, large numbers, spatial, temporal

INTRODUCTION

Geologic time is a fundamental idea that undergirds many geoscience concepts (Kortz and Murray, 2009), yet it is problematic for students. Undoubtedly, one reason for this is the size of the geologic timescale. Echoing previous national standards efforts, *A Framework for K–12 Science Education: Practices, Cross-Cutting Concepts, and Core Ideas* identifies scale, proportion, and quantity, which includes temporal as well as spatial scales, as cross-cutting concepts that impact all science disciplines, including the geosciences (National Research Council, 2011). Relationships among variables often change with scale. Some natural processes are not directly observable because they occur at spatial scales that are too large (or small) and/or temporal scales that are too long (or short). Geologic events and process that occur across many temporal and spatial scales are included in introductory university geoscience courses, as well as in K–12 science courses. Learners must understand something about the spatial and temporal scales at which geologic processes operate in order to develop scientific ideas about them.

There are two main components to an understanding of temporal scale across geologic time. The first deals with temporal **succession** or the capacity to place geologic (and biologic) events on an immensely large scale in relative and absolute positions (e.g., the extinction of dinosaurs preceded the appearance of hominids [relative] and occurred 65 Ma [absolute]). The second concerns temporal **duration** or the time period required for a geologic process or event to occur (e.g., the assemblage of Pangea took place over ~100 My).

Prior geologic time conception studies provide data and anecdotal suggestions regarding factors that might cause both succession and duration to be troublesome concepts for geoscience novices. Most research has explicitly focused on how learners understand temporal succession and often asks students to place geologic and biologic events in relative and/or absolute temporal order (Trend, 1998, 2000, 2001; Libarkin et al., 2007; Catley and Novick, 2009). Students and their teachers commonly place events in accurate relative order, thus being able to place the extinction of dinosaurs before the appearance of hominids. They have difficulty with more absolute measures (like 65 Ma) or determining how temporally close two events are to each other in both relative and absolute terms (Trend, 1998, 2001; Catley and Novick, 2009). Temporal compression of events farther back in time is common (Catley and Novick, 2009). Strategies to improve students' understanding of temporal succession typically focus on increasing their geoscience content knowledge (GCK), developing collective landmarks, or connecting the geologic timescale to their personal lives (Ritger and Cummins, 1991; Delgado, 2013). Intervention studies generally report modest increases in achievement pre- to postinstruction (Dahl et al., 2005; Teed and Slattery, 2011). While any improvement is positive, the fact that gains are not as great as might be desired suggests that there may be other factors at work. I will return to what those factors might be shortly.

Studies dealing explicitly with how students reason about geologic processes or how they understand their durations are fewer in number. Students often seriously underestimate the time required for a process such as folding and often attribute it to a catastrophic event (Hidalgo and Otero, 2004). These authors do not report whether students assigned a particular temporal interval to the fold's

Received 18 September 2012; revised 11 March 2013; accepted 1 June 2013; published online 27 August 2013.

[1]Childhood Education, Literacy, & TESOL, College of Education and Human Services, University of North Florida, 1 UNF Drive, Jacksonville, Florida 32224, USA.
[a]Author to whom correspondence should be addressed. Electronic mail: k.check@unf.edu. Tel.: 904-620-2794. Fax: 904-620-1025

formation, only that they did not ascribe its origin to processes that occur slowly over long periods of time.

Dodick and Orion (2003) investigated students' temporal reasoning related to the application of principles of relative dating, or relative temporal succession, but their work is also germane to a discussion of durations of geologic processes and events. In one task, they showed adolescents a line drawing of an outcrop containing five layers of rock, all of which were the same thickness. Layers one, three, and five were sedimentary layers, each containing a picture of a fossil. Layers two and four were igneous layers labeled 100 My and 90 My, respectively. Students were asked to determine the absolute ages of the fossil-bearing sedimentary layers. Overwhelmingly, students assumed all layers represented equal time periods (10 My) rather than realizing that the igneous layers provide boundaries for the dating of the sedimentary layers. The focus of the research question was on the age of the strata rather than the time required for its deposition and lithification. Nonetheless, as the authors state, responses could be taken as evidence that students in this study viewed spatial size and temporal duration as directly proportional (Dodick and Orion, 2003, p. 427).

In another task, the authors showed 52 11th–12th graders three drawings of a pair of outcrops in which thickness of strata and number of layers in the outcrops varied. Students were asked to determine which of the outcrops was older, if possible. No information about depositional rates was provided, so the correct answer was that relative age could not be determined from the drawing. Roughly one-third of participants answered each question correctly, and a higher percentage chose an incorrect option in each scenario. When number of layers varied but overall size was held constant, the outcrop with more strata was most frequently judged to be older. When thickness of strata was held constant but overall size varied, the taller outcrop was judged to be older. In both cases, spatial size or number was viewed as a marker and could be said to illustrate a "More A, More B" strategy (Stavy and Tirosh, 2000). Like the previous task, these results suggest that students view spatial size and time as directly proportional when determining age. Again, the item's focus was on relative age rather than the time required for the formation of sedimentary strata, so it is not clear if students hold the same view when judging a process's duration.

Lee et al. (2011) investigated students' ideas about the durations of a variety of processes/events across science disciplines. They found college students do well estimating durations for phenomena occurring on human timescales but often underestimate the durations of events/processes that require very long temporal periods while overestimating the durations of those that happen very quickly. Moreover, attempts to estimate temporal magnitude via bootstrapping were only somewhat more successful than guessing. As was true for all other studies discussed thus far, temporal periods were linked to specific events, and thus it is difficult to ascertain if students had poor conceptions of the time periods themselves, limited knowledge of the events, or both.

There is a need to build upon the very small body of research into geoscience students' ideas about temporal duration and explore whether those ideas are influenced by factors other than their GCK. Helping students develop understanding about rates and durations of geoscience processes and events is at least as critical as placing events on the geologic timescale. In fact, the development of accurate concepts about temporal duration may be the most important idea for geoscience novices to acquire (McPhee, 1981). The location of specific events on the geologic timescale only makes sense if one is able to conceive of temporal durations for geologic processes that require time periods far outside the bounds of human experience.

Several of the authors cited earlier herein mention factors that may influence how geoscience students acquire notions about temporal succession and duration. A tendency to equate spatial size and duration has already been noted (Dodick and Orion, 2003). Trend (1998, 2001), along with others (Dodick and Orion, 2006), alludes to the confounding role played by a person's understanding of large numbers on their ideas about temporal succession on the geologic timescale. Trend proposes that students do not sufficiently discriminate numbers in the thousands, millions, or billions, although this was not a variable of interest in his research. Two strands of research in cognitive science provide a framework in which to consider the ways in which these two variables might be influential. Together, they suggest ways to systematically explore whether ideas about spatial size and large numbers impact students' concepts about duration of geoscience phenomena.

Using Spatial Size to Infer Duration

Cognitive scientists have documented a close connection between spatial and temporal reasoning over short time periods, with the former influencing the latter (Piaget, 1969; Friedman, 1992; Boroditsky, 2000; Boroditsky and Ramscar, 2002; Gentner et al., 2002; Casasanto and Boroditsky, 2008; Casasanto et al., 2010). Specifically, there seems to be an inherent human bias to view space and time as covarying, thus larger events or processes require more time than smaller ones. When rates are held constant, the assumption is valid, but when they differ, it is not. Casasanto et al. (2010) argue that space serves as a marker for time because space is more directly experienced and durable, while time is abstract. The contexts for those studies have been non-geological; thus, at present, there is only a hunch that their findings apply to reasoning about geoscience phenomena. The importance of spatial reasoning to many areas of geoscience, such as the visualization of three-dimensional (3D) structures from two-dimensional (2D) drawings or the use of observations of several small outcrops to infer the nature of the larger formation of which they are a part, is well-recognized (Orion and Ault, 2007; Kastens et al., 2009; Ozdemir, 2010; Liben et al., 2011). In fact, Dodick and Orion (2003) found a correlation between scores on a spatial visualization task that assessed students' ability to visually manipulate 2D and 3D objects and their ability to correlate fossiliferous strata. The connection between the types of visual-spatial skills often associated with geoscience and the capacity to use principles of relative dating to determine temporal succession is important. Yet, it does not address the question of whether students view space and time as directly proportional when inferring temporal duration. Geoscientists regularly make use of durable, spatial information such as sedimentary strata in conjunction with other GCK such as depositional rates to infer temporal duration. The question is how do geoscience novices with more limited GCK view the relationship between spatial size and duration?

Ideas About Large Numbers

The temporal compression that occurs when estimating long time periods associated with geologic time (Catley and Novick, 2009; Lee et al., 2011) also characterizes young learners as they develop number sense for numbers of increasingly larger ranges. Some researchers speak of a "mental number line," which is an inherently spatial mapping of cardinal numbers (Dehaene et al., 2008; Izard and Dehaene, 2008; Kadosh et al., 2008). When children are familiar with numbers in a particular range (e.g., 0–100), they map those numbers in a linear fashion on a number line with equal intervals between adjacent whole numbers (Siegler et al., 2009). When mapping numbers in unfamiliar ranges, they tend to compress the relative distance between a pair of numbers farther away from 0 and exaggerate the difference between numbers closer to 0. As a result, larger quantities require a proportionally greater distance between them to be discriminable (Dehaene et al., 2008). Many see this as evidence of a logarithmic to linear shift in how learners perceive numbers, a perspective that changes with age and education (Siegler and Opfer, 2003; Booth and Siegler, 2006; Dehaene et al., 2008; Siegler et al., 2009; Thompson and Opfer, 2010). Of course, a true logarithmic scale is a highly sophisticated way to manageably represent quantities across many orders of magnitude on the same scale (Swarat et al., 2011). The scale children produce is more intuitive and might alternatively be described as a decreasing interval scale. Furthermore, it could be argued that the ability to create and interpret a true logarithmic scale is predicated on an understanding of the linear, proportional relationships among the quantities being considered. Absent such understanding, equal intervals on the scale are incorrectly interpreted to indicate equal absolute differences instead of equal ratios (Confrey, 1991).

An alternate view states that number line tasks are primarily proportional judgment tasks in which any number must be considered in light of the entire scale (Barth and Paladino, 2011). In this view, people underestimate and overestimate proportions in predictable ways depending upon whether only the largest value is considered when estimating or if intermediate reference points are also used to make proportion judgments. In either case, estimation errors can be explained on the basis of proportional reasoning without reference to logarithmic or linear mapping. Whether nonlinear number mapping is better explained by a logarithmic to linear shift or proportional reasoning, the task is practically impossible if there is minimal understanding of the magnitude of the largest value (e.g., 100 or 1,000). The parallels to geologic time periods, in relation to either duration or succession, are perhaps obvious. The scale on which the time period is being considered is critical when judging its duration relative to other time periods.

The studies just described dealt with cardinal numbers, not temporal periods, and often with quantities that are quite small in geologic terms (usually 1,000 or less; Dehaene et al. [2008] being a notable exception). Millions and billions of anything are generally outside the realm of human experience. If geoscience novices have unclear notions of how quantities of those magnitudes differ from each other, the difference between 1 million and 1 billion years may **seem** smaller to them than is the case. This could explain some of the temporal compression seen in earlier geologic time conception studies. Some students may not possess knowledge of large numbers and the proportional relationships among them robust enough to deal with long temporal durations and certainly not with the large temporal periods of the geologic timescale. To date, there has been limited data that specifically relate learners' ideas about large numbers to their temporal understanding of geologic phenomena (Confrey, 1991; Cheek, 2012).

Research Questions

This study builds upon and extends the studies discussed here by exploring the influence of ideas about spatial size and large numbers as they relate to temporal durations of a variety of geologic processes across the temporal scales commonly encountered in introductory geoscience courses. In doing so, the study is not focused on students' GCK. No geoscientist would deny the critical role played by GCK in learners' conceptual understanding. Previous geoscience research in this area has linked temporal markers to specific geoscience events (Trend, 1998, 2000, 2001; Dodick and Orion, 2003; Hidalgo and Otero, 2004; Libarkin et al., 2007). Obviously, learners must be able to make those connections. The difficulty is that if research questions are only explored within a geoscience context, we are left with a conundrum. We can say, as has already been said, that many introductory geoscience students have relatively inadequate conceptions of long time periods, but we cannot say why that is the case. We may postulate that it is because they lack meaningful geoscience referents (GCK), but we cannot be sure that other factors are not also at work.

Two questions are investigated:

(1) Do geoscience novices equate spatial size with duration for geologic events and processes across a variety of temporal scales?

(2) Do geoscience novices display robust understanding of the relative sizes of long temporal periods (up to 100 million years)?

The two questions are investigated simultaneously because they are closely related. As indicated already, both have been noted by other researchers as possible confounding variables in students' conceptions about geologic time. Some view space, time, and number as part of a larger system of magnitude, with each exerting similar influence upon the others (Walsh, 2003), though others argue (Casasanto et al., 2010) that ideas about time and number are both influenced by spatial thinking. In either case, there is sufficient overlap to warrant their joint consideration in the same study.

This study is not focused on conceptual change. Rather, its goal is to ascertain whether ideas about the relationship between space and time along with large numbers may account for introductory geoscience students' ideas about temporal duration across temporal scales. Results of this work may suggest avenues for conceptual change research.

METHODS

Participants

Seventeen university undergraduates (12 females and five males, median age 21 y) enrolled in an introductory geoscience course taught by the investigator at a small religiously affiliated U.S. liberal arts college comprised the nonrepresentative sample. The school has no official policy on young-Earth creationism, though it is affiliated with a religiously conservative denomination. An announcement was made in class midway through the semester soliciting study volunteers who were asked to privately contact the instructor via electronic mail to indicate their willingness to participate. All gave informed consent and received modest extra credit points on a course exam for their participation (one of several ways to receive the same number of extra credit points on a single course exam).

Data were collected during the semester, but analysis was not begun until after all course grades were submitted to the registrar. Mean grade of participants across four exams prior to inclusion of extra credit points was 68 (standard deviation [SD] = 12). Interviews occurred near the end of the course after units on plate tectonics and tectonic processes, the rock cycle, relative and absolute dating methods, and surface processes. Ten of the 17 participants reported prior geoscience coursework in either middle school or high school. An additional student took an oceanography course at a community college. Six participants said they did not recall studying geoscience prior to the current semester.

Task-Based Interviews

Volunteers participated in individual, semistructured task-based interviews. Cognitive interviewing has a long history in conceptions research (Piaget, 1969; Posner and Gertzog, 1982). Tasks around which to situate interview questions were employed to provide a problem-solving context for students' explanations of their reasoning strategies and to investigate the strength of responses by evincing response consistency across tasks. Follow-up questions were asked to elicit additional information. The limitations of cognitive interviewing are well documented (Posner and Gertzog, 1982; Marin, 2004; Myers and Newman, 2007). Samples of student responses are provided to enable the reader to judge the trustworthiness and reliability of the findings. The supplemental materials contain a sample interview script (Supplement 1; available at http://dx.doi.org/10.5408/12-365s1). The order of presentation of animations was counterbalanced, so the order of questions in that portion varied across participants. There was no time limit, but most interviews lasted approximately 30 min. They were audio-recorded and fully transcribed.

Analysis of interview responses as well as artifacts produced proceeded from a description as interpretation framework using a cross-person approach (Yin, 2011). Initial and axial coding schemes were used to disassemble and reassemble data (Charmaz, 2006). Some descriptive statistics, such as means and percentages were calculated, but data analysis was primarily qualitative. Interview transcripts were analyzed to determine themes in responses (Creswell, 2009).

Description of the Tasks

Three of the interview tasks were purposefully devoid of geoscience content because students' GCK was not a variable of interest. Pre- and postassessment of student understanding about temporal duration or geoscience concepts more broadly was not conducted. The reason for this decision was described in the Introduction. Two tasks were placed within a geoscience context in order to compare responses on those items with the nongeoscience ones. See Table I for a list of tasks and brief descriptions about them.

Question 1: Spatial Size and Duration (Tasks 1 and 2)

Tasks 1 and 2 explored how students viewed the relationship between spatial size and temporal duration. For task 1, students were shown a line drawing of an outcrop containing Phanerozoic sedimentary strata and Precambrian base rocks (with no identifying information on the drawing to indicate that it was the Grand Canyon). Participants were asked to compare the relative length of time required for a specific sedimentary layer and a thinner one above it to form (response choices: one longer than the other, both same amount of time, or cannot be determined from drawing). Specific rock types were not identified for participants. The correct response is that the answer cannot be determined from the drawing alone, as information about the depositional environment would be required. This task differed from the one used by Dodick and Orion (2003), which investigated students' ideas about relative age (succession). The present task was concerned with their thoughts about temporal duration.

Students next watched three computer animations of colored layers filling under various conditions (task 2). Animations were designed with the help of a technology specialist at the college. Still shots of the animations are available in the supplemental materials (Supplement 2; available at http://dx.doi.org/10.5408/12-365s2). The only connection these animations had to geoscience is that they were horizontal layers that appeared from bottom to top consistent with superposition. They were not designed to simulate deposition, but rather to explore whether participants would equate spatial size with temporal duration over short time periods for a novel task. Filling durations ranged from 4 to 10 s. A timer in the upper-right corner of the screen ran continuously for the duration of each animation, but its presence was not pointed out to students. The use of an instrument to measure the passage of time has precedent in temporal duration research, and children, at least, do not always deem it to be a consistent measure of time (Piaget, 1969).

After viewing each animation as often as they wished, participants compared the filling durations of two pairs of layers. All were reminded they could watch the animations again, if they wished. Questions followed the same pattern across animations. Question 1 asked which of two layers took longer to fill, and the second asked which of two layers filled more quickly. Order of the animations was counterbalanced. In Animation 1, all layers were the same thickness but filled at different rates, and had different durations. Both thickness and rate varied in Animation 2, but all durations were the same. Layers in Animation 3 varied on both dimensions, but the variability was greater than for Animation 2. Some pairs of layers in Animation 3 had either the same duration but different thicknesses as in Animation 2 or the same thickness but different durations as in Animation 1. A small pad of paper and a pencil were available for interviewees' use if they wished to record starting and ending times for each layer.

After watching the animations, students were again shown the task 1 line drawing. The initial question was repeated, and they were asked to answer in light of the animations. Responses to the first time the question was asked could reasonably be construed to only be accessing

TABLE I: Description of tasks used to investigate each research question.

Research Question	Task	Description
A. Do geoscience novices equate spatial size with duration for geologic events across a variety of temporal scales?	1. Sedimentary strata line drawing	Students compared the relative durations for the formation of two adjacent sedimentary layers. This was done twice—prior to and after completing task two.
	2. Duration animations	Students watched two animations of colored layers filling in real time with different durational periods. Students then compared the relative durations of two layers.
B. Do geoscience novices display robust understanding of the relative sizes of large temporal periods (up to 100 million)?	3. Time line 1	Students placed four time periods on a line based upon their ideas about the proportional relationships among the temporal durations.
	4. Time line 2	Students placed seven time periods on a line based upon their ideas about the proportional relationships among the temporal durations. Time periods corresponded to scientifically accepted time periods for the events in task 5 rounded to the nearest power of ten.
	5. Geoscience Events Time line (GET)	Students placed seven geoscience events on a line based upon their ideas about the proportional relationships among the events' durations.

students' GCK. When the question was re-asked after watching the animations, information from them about possible relationships among rate, size, and duration was available to students.

Question 2: Ideas About Long Temporal Periods (Tasks 3–5)

Two of these tasks dealt with large numbers in a nongeologic context, while a third had a geologic component. Students completed two numeric time lines (tasks 3 and 4) with time periods up to 100 million years, the approximate duration of the breakup of the supercontinent Pangea, a geologic process commonly discussed in introductory geoscience courses. While 100 million years is only a fraction of Earth's history, it extends most of the number studies described in the Introduction by many orders of magnitude and connects them to temporal durations for a number of geologic processes. Time periods for tasks 3 and 4 are listed in Table II. For each time line, students were given a 40-cm-long strip of paper with a line drawn horizontally across the middle and a list of times, which they were asked to place in order on the line "based upon how long the time periods take to pass in proportion to each other" for each time line. Times for task 3 were written on separate index cards and placed in ascending order from left to right above the time line. Times for task 4 were listed on a single sheet of paper, each next to a corresponding letter to save time for

TABLE II: Stimulus items for time lines 1 and 2.

Task 3: Time Line 1	Task 4: Time Line 2
1,000 y	1 min
100,000 y	1 d
1,000,000 y	1 mo
100,000,000 y	1 y
	10,000 y
	10,000,000 y
	100,000,000 y

participants and make discrimination of the locations of the seven time periods easier to discern. To determine the times for task 4, the investigator took events used for task 5, the Geoscience Events Time Line (described below and in Table I), and rounded their durations to the nearest power of 10. Numbers were rounded to make proportional reasoning calculations easier by enabling participants to rely solely on base-ten proportional relationships rather than less easily manipulated ratios.

Prior to the interviews, the investigator prepared a time line with 1 s, 1 min, and 1 d in which times were placed proportionally based upon their durations to provide assistance for students who might be confused by the task. It was not routinely shown to all participants lest it be treated as a template to follow that would eliminate the need to reason about the relative sizes of the numbers involved. One student asked whether it mattered if she began on the left or right. A second asked if she only needed to place times in order. Both were shown the sample time line with an accompanying verbal explanation as to why times were placed where they were.

Time lines were independently sorted by the investigator and another scorer, who was not otherwise involved with the study, on the basis of predetermined criteria focused on degree of linear, proportional placement of time periods across the line. As described above, linear mapping of cardinal numbers has been seen as evidence of number sense for numbers in a particular range (Siegler and Opfer, 2003; Dehaene et al., 2008; Siegler et al., 2009; Thompson and Siegler, 2010; Barth and Paladino, 2011).

Initial interscorer agreement was 88% (15/17) for task 3 and 94% (16/17) for task 4. Discrepancies between scorers were reconciled through discussion and joint agreement, since they only differed by one category. After initial sorting, the investigator used interview transcripts to ascertain if someone's explanation suggested a better understanding of large numbers than was evident from the time lines alone. A nonlinear time line could result from a number of factors, some unrelated to ideas about numbers. Thus, the artifact alone provides insufficient evidence of a poor understanding of large numbers. One time line was moved to a higher

TABLE III: Stimulus items for task 5: Geoscience events time line (GET).

A.	The Earth spinning around once
B.	How long most coral reefs have been growing
C.	The breakup of the supercontinent Pangea
D.	The Earth going around the Sun once
E.	The Moon going around the Earth once
F.	The carving of the Grand Canyon by the Colorado River
G.	The amount of time the ground shakes during an earthquake

category based upon the person's verbal description of it. None was moved to a lower category.

Task 5 linked the temporal periods used in task 4 to specific geoscience events. Students placed the events listed in Table III on a Geoscience Events Time Line (GET). Length of time line and setup of paper were identical to tasks 3 and 4 with one exception. The words "shortest" and "longest" appeared on the far left and right sides of the line, respectively. Students were instructed to place events on the line in relative order based upon their temporal durations. A similar design employed by Libarkin et al. (2007) dealt with temporal succession, while this one focused on the relative durations of the events. Events taking around the same amount of time were to be put close together, while those for which the durations were very different were to be spaced far apart. Temporal periods of some events on the GET are not generally assessed in introductory geoscience courses, so correct linear placement of events was not analyzed for the GET as it was for tasks 3 and 4. Placements were compared with those of time periods on task 4 to determine the congruity between ideas about large numbers and the durations of specific geoscience events or processes. If students performed well on the numeric time lines but not on the GET, that would suggest that they are able to deal with long temporal periods, but may not have sufficient GCK to associate specific processes with those time periods. Conversely, students who perform poorly on both the numeric time lines and the GET may do so because they lack solid understanding of large numbers and the proportional relationships among them, and/or they are unfamiliar with the geoscience events involved. The nature of this instrument does not provide answers to the relative importance of those factors in students' responses. Nonetheless, it indicates that helping those individuals better understand geologic time will require something more than just increasing their GCK.

RESULTS

Results for tasks related to each research question are reported and briefly discussed in turn in the order in which they were completed with the exception of the fact that task 2 is discussed before task 1.

Question 1: Tasks 1–2

Animations (Task 2)

Inferring temporal duration of deposition of geologic strata in the field requires significant GCK. Animation layers formed in real time, so the ability to complete the task successfully could not be attributed to an individual's GCK. While this task might be deemed a simple one for adults, students did not consistently judge relative durations correctly. Sixty-five percent of interviewees answered three or four questions out of six correctly across animations. They judged duration accurately slightly more than half the time (56%) when layers were the same thickness but filled at different rates, and, thus had different durations (Animation 1). Accuracy was lowest for Animation 2 (29%), where layers were of differing thicknesses but filled in the same time period, a finding that accords with an earlier, similar study (Cheek, 2011). Students were most successful in judging durations on Animation 3 (79%). Higher accuracy for Animation 3 was true regardless of the order in which animations were seen. One pair of layers compared were of the same thickness; thus all strategies other than equating size with duration should have been successful. The variability in rate between the second pair of layers was greater than for any pairs across animations, thereby making discrimination easier.

Everyone chose to watch at least one animation a second time, but the number of times a person watched the animations appeared to have no effect on accuracy of responses. Two students used the paper and pencil to record the filling durations of layers. The first was correct 50% of the time, and the second was correct 83% of the time. A third student asked to use paper and pencil, did not actually write anything down, and was correct 33% of the time.

Sample responses with initial and category codes generated from those responses can be found in Appendix 1. Reasons for duration judgments varied across animations and by participant. Table IV lists results for both tasks 1 and 2 along with the category to which a participant's explanation belonged. The fact that students cited different reasons across animations is consistent with the ways in which the animations varied. Equating size with duration was an insufficient strategy for all animations. Co-consideration of rate and size or the use of iterated, equal temporal units should have produced correct responses in all cases. Inverse proportionality of rate and duration should have produced correct responses for Animation 1, since all layers were the same thickness in Animation 1 and filled at different rates. It would have also worked for one of the Animation 3 questions. Inverse proportionality would have been insufficient for Animation 2 and the other Animation 3 question, since size varied, and rate by itself can only be used to judge duration if size is held constant.

Results were at chance level for Animation 1 (four correct, four incorrect) when using inverse proportionality of rate and duration (which should have produced correct responses). That strategy was cited alone 11 times (three correct) for Animation 2, but it should not have given a correct answer if it was the only factor considered, since size of layers varied. Equating size with duration (where thickness of layers differed) was more common for Animation 2 (four times) than either of the others (once each). Size was held constant for Animation 1 and one question for Animation 3, which made it less likely to be viewed as a salient variable. It cannot be unequivocally stated that students who said they used size to judge duration or those who mentioned inverse proportionality of rate and duration were **not** considering the two factors in

TABLE IV: Results for tasks 1 and 2 by participants.

Participant	Task 1 Before Animations	Task 2			Task 1 After Animations
		Animation 1 (Number Correct)[3]	Animation 2 (Number Correct)[3]	Animation 3 (Number Correct)[3]	
F1	Larger[1]	$0^{C,C}$	$0^{D,A}$	$2^{D,D}$	Don't know
F2	Larger	$2^{D,D}$	$0^{A,A}$	$2^{B,B}$	Don't know[1]
F3	Larger	$1^{D,D}$	$0^{D,D}$	$2^{D,D}$	Don't know
F4	Larger	$1^{C,B}$	$0^{B,B}$	$2^{B,B}$	Don't know
F5	Larger	$0^{B,B}$	$0^{B,B}$	$1^{B,B}$	Smaller
F6	Larger	$1^{D,D}$	$2^{D,D}$	$2^{D,D}$	Don't know[1]
F7	Larger	$0^{E,E}$	$1^{B,B}$	$1^{B,B}$	Smaller
F8	Larger	$1^{B,B}$	$1^{E,E}$	$2^{B,E}$	Smaller
F9	Larger	$2^{D,D}$	$0^{D,B}$	$2^{B,B\&D}$	Don't know
F10	Larger[1]	$0^{D,D}$	$1^{B\&D,A\&B\&D}$	$2^{B,B}$	Don't know[1]
F11	Larger	$2^{C,B}$	$0^{C,C}$	$0^{C,E}$	Larger
F12	Larger	$1^{D,B}$	$1^{B,E}$	$2^{D,D}$	Don't know
M1	Larger	$2^{C,C}$	$0^{C,C}$	$2^{C,B}$	Smaller[1]
M2	Smaller[2]	$2^{E,E}$	$1^{B,B}$	$0^{B,B}$	Smaller
M3	Don't know[1]	$2^{D,D}$	$2^{B\&D,B\&D}$	$2^{D,D}$	Don't know
M4	Don't know	$1^{A,B}$	$0^{A,E}$	$2^{A,B}$	Smaller
M5	Don't know[2]	$1^{D,A\&D}$	$1^{D,B}$	$1^{D,B}$	Larger

[1]Geoscience vocabulary germane to task 1, like deposition, sediments, and depositional environment, whether directly stated or by use of a definition.
[2]Geoscience vocabulary that is unrelated to task 1, like metamorphosis or anticline.
[3]Participant codes were created by assigning each student a number with a prefix indicating gender. Reasons cited correspond to category codes from Appendix 1. Category codes are listed for questions one and two, separated by a comma. When category codes are joined by an "&," they were mentioned together for the same question.

tandem, only that they did not mention the two together. When participants specifically said they considered the inverse proportionality of rate and duration in conjunction with size of the layers for the three questions in which rate, size, and duration all varied, they were correct 75% of the time.

The most reliable way to judge duration would be to use standard iterated units. It was moderately successful when judging durations for Animations 1, either alone or in conjunction with another strategy (63%), and highly successful for Animation 3 (100%), but only somewhat better than chance for Animation 2 (55%). Casasanto and Boroditsky (2008) found college students do not always judge short duration correctly, but it is interesting that the timer did not result in greater accuracy here. Any mathematical computation should have been well within the capability of a college student; animations could be rewound, and paper and a pencil were available to record times. Students may have been insufficiently motivated to pay careful attention to the timer. Yet, one woman used the timer for all three animations and wrote down starting and ending times for all layers, but was only correct half the time. It is hard to invoke lack of motivation as a reason for her difficulty.

The relatively low accuracy when using the timer mirrors Piaget's results with children (Piaget, 1969), which he attributed to their greater reliance on perceptual information as opposed to an objective temporal unit. At least three students appeared to experience some conflict between the size of the layers and filling rates or use of iterated units. The following interchange with one woman about Animation 2 illustrates this point.

> *F10: I think the brown layer would fill up first, but again they did it at the same time, so I'm confused.*
>
> *Interviewer: Tell me about why you're confused.*
>
> *F10: Well, it went up at the same time. It seemed like it was going at the same rate, but other colors immediately went up, then the other ones didn't but they had the same amount of time. It didn't look like it was going any faster.*
>
> When asked to clarify, she said:
>
> *Well, the brown would fill up first, but it shouldn't be happening. It should be coming at the same time because it was moving at the same rate, or it looked like it was moving at the same rate, but there's clearly more brown than there is pink.*

This woman mentioned the timer earlier than what is reported here, so she had an objective measure of duration, but she did not appear to see it as the most reliable way to judge duration. She mentions rate differences but also describes filling rates as constant (they weren't). Although she doesn't specifically articulate the size difference between the layers until near the end of the interchange, her mention of them may suggest that she viewed size as an indicator of duration.

Sedimentary Strata Line Drawing (Task 1)

Table IV lists responses by participant along with accompanying reasons for the task 1 question before and after the animations. The propensity to equate spatial size with duration was more common in students' responses here than in the animations. Prior to watching the animations, 13 of the 17 interviewees said the thicker layer would take longer to form simply because it was thicker.

> *I'm going to go with thicker, it took longer 'cause when I'm cooking something in the oven if the cake batter's thicker it takes longer. (F10)*

The animations were not specifically designed as a teaching intervention, but 15 students changed their answer on task 1 after watching them, though not all changed to the correct one. One of the two students who did not change his answer was correct prior to watching the animations. The second student answered only one of the animation questions correctly. She cited intuitive perception as the reason for her animation responses and, even after probing questions were asked, would not elaborate further. After the animations she continued to assert that a thicker sedimentary layer had a longer depositional period than a thinner one. In response to a follow-up question, she acknowledged that some thin layers filled more slowly than thicker ones in the animations, but insisted thinner sedimentary strata would require shorter depositional periods **because** they were thinner even if they filled at slower rates.

After watching the animations, eight students said they would need more information to determine which of the two sedimentary layers took longer to form. Five of those said size and duration did not covary in the animations, while three said they lacked information about depositional rates, the salient factor.

Geoscience vocabulary was not introduced by the investigator during the study, and participants' spontaneous use of terms differed before and after the animations. Even though they had already been taught about relative dating and sedimentary processes in the course, students rarely used relevant terms, or any geoscience terms for that matter, **prior to** viewing the animations. "Anticline," metamorphosis," and "sediment" were mentioned by one student each. Three people used "erosion" (two used the term specifically, and in another case it was implied). One thought possible erosion was indicative of relative age, i.e., the thinner stratum on top was older than the thicker, underlying layer because it may have been eroded. The 23-y-old student who answered correctly **both** before **and** after the animations was the sole individual to state that the thickness of exposed rock layers reflects not only the depositional period but also what happened since deposition occurred before the layer above it was laid down.

> *Layer 3 is larger than layer 4, but I wouldn't necessarily say that it took longer time to build because it, obviously they're different types of rocks. So with layer 3 being laid down over time you automatically have to think of erosion and all that kind of stuff as well....even though it might be the same amount of time that 3 and 4 were both made, 4 might be smaller because it breaks down more, it breaks down faster. (M3)*

Even though the animations were nongeologic, they appeared to activate prior GCK. **After** watching them, three people used the word "deposition" or some variant in their explanation. Of the eight people who switched answers from incorrect to correct, five specifically cited "rate" in their explanation, though not all coupled it with deposition. Minimally, the animations appeared to stimulate thinking that rate must be considered along with size when judging durations and that a thicker sedimentary layer does not automatically signal a longer depositional period than a thinner layer. When asked how the animations helped her answer the question, one woman said,

> *We just can't make the presupposition that every single sediment will settle at the exact rate of time, even though they look the same in that picture. (F6)*

Following the animations, five individuals said the thinner layer took longer to be deposited and cited the animations, even though thinner layers did not always fill more slowly. Inferring why students reached this conclusion involves some conjecture. One woman expressed the view that what she had seen in the animations was counterintuitive. That may have made it more noteworthy for her.

> *You would think that small amounts would be built up faster, but they [larger layers] took less time. (F5)*

As is typical for drawings like the one used for task 1, rock types were indicated by patterns that enable geoscientists to see at a glance which rock types make up the formation. Interviewees were told that the patterns would not help them answer the question; nonetheless, the patterns were distractors for nine students, particularly the first time the question was asked. Five thought the "brick" design of one stratum indicated multiple smaller layers within the larger layer and, thus, a longer depositional period. Two people described the thinner layer as either more broken up or containing particles because it appeared to be composed of small "pieces." This raises concerns about the inferences geoscience novices may be making from geologic diagrams in textbooks and other instructional materials and the misconceptions about rock formation that could result.

Question 2: Tasks 3–5

Time Lines 1 and 2 (Tasks 3 and 4)

Samples of time lines scored as "Correct," "Partial," and "Incorrect" along with explanations for scoring can be found in the supplemental materials (Supplement 3; available at http://dx.doi.org/10.5408/12-365s3). After preliminary sorting as described in the Methods section, participants were placed into two groups: linear, proportional (LP); and intuitive, nonlinear (IN).

Those who scored correct or partially correct on both time lines were classified as LP, while those who scored incorrect on one or both lines were classified as IN. The dichotomous grouping undoubtedly resulted in some people being classified as LP whose inclusion in the category is questionable. Any systematic error that resulted would tend to refute the study's claims. Table V lists students by their classifications and indicates whether they described their placement of numbers for tasks 3 and 4 in qualitative or

TABLE V: Participant category placement along with explanations for tasks 3 and 4.

Participant	Category Placement	Time Line 1 Explanations (Task 3)[2]	Confusion about Size of Numbers Explicitly Expressed (Time Line 1)	Time Line 2 Explanations (Task 4)[2]
F1	IN	Qualitative		Qualitative
F2	IN	Qualitative		Quantitative & qualitative
F3[1]	IN	Qualitative & quantitative (after probe)	X	Qualitative & quantitative (after probe)
F4	IN	Qualitative		Qualitative
F5	IN	Quantitative & qualitative	X	Quantitative & qualitative
F6	LP	Qualitative		Qualitative
F7	LP	Quantitative		Qualitative
F8[1]	IN	Quantitative & qualitative		Quantitative & qualitative
F9	IN	Quantitative	X	Quantitative & qualitative
F10	LP	Qualitative		Qualitative
F11	IN	Qualitative	X	Qualitative
F12	IN	Qualitative		Qualitative
M1	LP	Quantitative		Qualitative
M2	IN	Quantitative		Qualitative
M3	LP	Quantitative		Qualitative & quantitative (after probe)
M4	IN	Qualitative	X	Qualitative
M5	IN	Quantitative		Qualitative

[1]Indicates those who were shown a sample time line.
[2]The order in which "Qualitative" and "Quantitative" appear for a participant in columns 3 and 5 indicates the order in which those descriptions were applied. In cases where a quantitative explanation was not spontaneously given, but only in response to a probe, that is indicated in parentheses.

quantitative terms. Examples of qualitative descriptions were:

- "The numbers gradually increase."
- "It was going to be a bigger amount to get to the next one."
- "I was thinking 1,000 years, then 100,000, then a million, then 100 million, so I was kind of like, I was going in order."

Students using quantitative descriptions attempted to describe mathematical relationships among the time periods, though they did not always do so accurately. Examples are:

- "Well, they're like 100 of each other or ten?"
- "99 million years is way more than 99 thousand."
- "A hundred thousand years is 100 of a thousand years."

Twelve interviewees were classified as IN. Half placed times fairly evenly across the line for task 3, as if all temporal periods were of similar length. The remaining six made some distinction among the spaces between the numbers, though only two placed 1 million years to the left of the line's midpoint. Five demonstrated some confusion about the relative sizes of the numbers. Three thought 100 thousand would be equivalent to 100 million because both are one hundred times their base unit (thousand and million, respectively), thereby creating a rudimentary log scale.

Examples of this type were determined to be rudimentary or intuitively logarithmic rather than true logarithmic scales. Not all adjacent pairs of numbers differed by a factor of 100, so there was at least some confusion about proportional relationships among quantities. If someone was attempting to create a genuine logarithmic scale, it should have been evident in the distance between 100,000 and 1,000,000 y, but that was not the case. This is not to imply that individuals in this group were completely unaware of proportional relationships between temporal periods, or that their time lines were identical. Half said time periods differed significantly from one another and often mentioned there was a difference between adjacent numbers. However, they were either unable to describe the differences quantitatively or demonstrated confusion about powers of ten.

Four students in this group placed times fairly evenly across the line for task 4 as they had done for task 3. Eight created **some** distinction between the size of temporal periods up to 1 y and those longer than that. There was considerable variation in where they placed 1 y. Its placement ranged from one-twentieth to one-fourth the total distance of the line from the point of origin. Four of the eight placed the three largest time periods (10,000; 10,000,000; and 100,000,000 y) equidistant from each other, while the others placed the two largest periods closer together than the fifth and sixth. Five described the quantitative relationship between at least one pair of time periods up to 1 y, but after that spoke only in vague, qualitative terms about the differences between larger periods (see Table V). They used changing units across the time line, but not in a way that would reasonably demonstrate proportional relationships among quantities or a logarithmic scale. Spaces between adjacent time periods up to 1 y were expanded, while those from 10,000 to 100 million years were compressed. In general, they appeared to use 1 y as an intermediate reference point to make proportion judgments by relating both smaller and larger numbers to it, almost as if they were trying to place two distinct scales on the same line (Barth and Paladino, 2011).

Five students created time lines that showed reasonable linear placement of time periods on the scale and were thus labeled as LP. Four of the five either described the tasks as mathematical and/or mentioned quantitative relationships. That did not mean that they accurately stated the proportional relationships between numbers correctly:

> *Yeah, I would break everything down under one hundred million. So 1 million of 100 million would be 10%. Then the 100,000 would be—is that 0.1%? Then the 1,000 would be 0.001% of the time. So that's how I'll break it down. So, out here on the far end would be 100 million—way out here. Everything else would be broke[n] down in here in the last 10%. (M3)*

LPs also overestimated the size of temporal periods up to 1 y on task 4, though they did so to a far lesser extent than INs. The farthest any LPs placed 1 y from the origin was one-tenth of the way across the line. Even though there was greater similarity in what the two groups drew on the second time line, LPs were more likely to describe their time lines in terms of proportional relationships than those of INs. Admittedly, task 4 was extremely difficult. Task constraints (size of paper; no gridlines or measuring tools) and/or the cognitive demands of placing time periods across so many orders of magnitude on the same line could account for overestimation of small quantities unless one was constructing a true logarithmic scale.

Four people mapped time periods right to left instead of left to right. This stands in contrast to cardinal number mapping studies that indicate English speakers consistently map numbers left to right (Petitto, 1990; Izard and Dehaene, 2008; Kadosh et al., 2008; Siegler et al., 2009). The difference in directional orientation is intriguing, but it may still reflect a common underlying left to right orientation for time as well as number. The fact that interviews were conducted by their geoscience instructor could have primed students to think about temporal periods as time periods that began in the past and continued to the present or future. If they perceive time as moving from left to right, it is reasonable to place longer temporal periods farther to the left, as they represent greater distance from the present. One student in an earlier, similar study exhibited the same tendency (Cheek, 2012).

In contrast to INs, all LPs began by placing 100,000,000 y first and then positioning other time periods in relation to it. INs usually began with the smallest time period and placed all others in order from smallest to largest, though one person placed 100,000,000 y immediately after 1 min. By changing the scale as they moved across the line, larger time periods were almost inevitably placed too close together. On that basis, it might be argued that differences between groups can be accounted for by problem-solving strategies alone, with no regard for ideas about long temporal periods. Conversely, perhaps better problem-solving strategies are used **if** a person has robust ideas about how the time periods are related to one another, specifically, their relationship to the largest time in the group. Those with greater mathematical understanding approach the task in a more systematic, mathematical way.

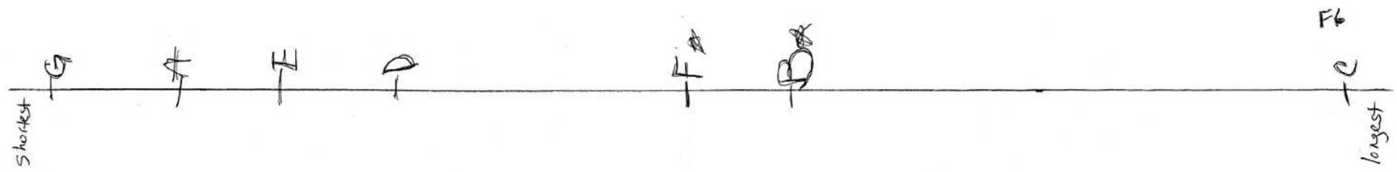
GET (Task 5)

The Geoscience Events Time Line (GET) added a GCK dimension to the duration time-line construction task. To draw a reasonably accurate time line, students needed correct ideas about proportional relationships among time periods coupled with specific content knowledge about the durations of those events (or the ability to extrapolate duration based upon other GCK). A nonlinear time line could result from either insufficient GCK about the event's scientifically accepted duration and/or an undifferentiated understanding of large numbers. Reasons for placement of events on the GET were analyzed qualitatively. Results from the GET were also compared with task 4. One student refused to complete the GET, saying she did not know the durations of any of the events. A second student's GET and descriptions were deemed unable to be analyzed. He did not settle on a clear response for six of the seven events, even after probing. There was sufficient concern about the credibility of his responses due to their shifting nature, so they were excluded from this portion of analysis.

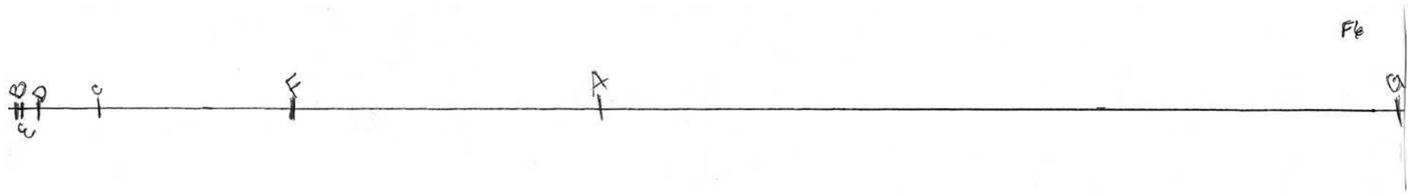
Students displayed relatively inaccurate ideas about the durations of these events, even those they would have likely encountered multiple times in school, such as the Moon's period of revolution. Six students placed events fairly evenly across the line. The others created some distinction between durations for events up to 1 y and those greater than that, as in Figure 1a, which was similar to what was observed on task 4. Arguably, lack of appropriate GCK could explain any difficulty LPs had with the task, since they appear to have had sufficient understanding of large numbers to cope with long temporal periods. Less can be said about those in the IN group. Their difficulty with the GET could be at least partially explained by their unsophisticated ideas about large numbers, which may or may not have been accompanied by inadequate GCK.

Table VI lists qualitative and quantitative temporal descriptions for the durations of events on the GET (task 5). Some imprecise language was used for durational periods of 1 y or less, but it was far more common when describing events with long durational periods. Even when prompted to offer specific numeric time periods, nine were unable to do so for more than one event, and only spoke in vague terms, as shown in column two of Table VI. The inability of participants to quantify what was meant by phrases such as "a really long time" is consistent with results for tasks 3 and 4. Two students said coral reefs have been growing since "creation," but neither would attach a specific time period to that statement.

When participants did name specific durations, time periods mentioned often differed significantly from scientific consensus. An examination of column three in Table VI shows that students' ideas about the duration of many geoscience events span several orders of magnitude. The only items on which there was full agreement, at least in quantitative terms, were Earth's rotation and revolutionary periods. There was some confusion about the difference



1a. GET



1b. Timeline 2

FIGURE 1: Comparison of GET (1a) and time line 2 (1b) for F6: 20-y-old female. Even though she said Earth's orbital period is 1 y, note that its placement (D on 5a) on the GET is much farther from the origin than its placement (C on 5b) on time line 2. Stars on the GET indicate two events this woman was unsure about.

between the "Earth spinning around once" and the "Earth going around the Sun once." The terms "revolution" and "rotation" were deliberately **not** used so as to mitigate such confusion, but it still occurred.

Equating size with duration was common. The depth of the Grand Canyon was cited by three individuals as evidence that its carving by the Colorado River took a long time. On the other hand, one student described the carving of the Grand Canyon as a potential natural disaster that could happen quickly. Two women tried to use information about annual rates—one of seafloor spreading and the other of coral growth—to infer durations, though neither was able to do successfully, as others have found (Lee et al., 2011).

Other alternative conceptions surfaced. A 20-y-old female said the Earth rotates on its axis while remaining in a stationary location in space. Five explicitly said the Moon orbits the Earth in a day. One person said the Earth and Moon rotate together, and therefore the Moon orbits the Earth in one day. Three people asked if one could ascertain the duration of a process if it was still occurring. This was

TABLE VI: All ideas expressed about durations of events on task 5: Geoscience events time line (GET).[1]

Event	Qualitative Duration Descriptions	Quantitative Duration Descriptions
Amount of time earth shakes during earthquake	Relatively short, not a super long time, not as long as the other ones, don't hear about those happening for very long	Seconds, minutes, 5 min, 10 min, <day, day, week
Earth spinning around once	No qualitative descriptions given	Day, 24 hours
Moon going around the Earth once	Long time, awhile	Day, 24 h, 27ish d, 30 d, monthish, months or years
Earth going around the Sun once	Decent amount of time, a long time	Year
How long most coral reefs have been growing	Since God created them, since creation, how old the Earth is, awhile, not that long, a lot of time, a long time, a really long time, a long, long time	Hundreds of years, thousands of years, maybe millions, few million years, 100 million years
Carving of the Grand Canyon by the Colorado River	Larger period of time, lots of years, awhile, a lot of time, really long time, a long time, not too, too long—just a natural disaster—could potentially go quickly, a longer time, didn't happen in a couple years, a very long time	Hundreds of years, thousands of years, maybe millions, 50–100 million years
Breakup of the supercontinent Pangea	Awhile, larger period of time, didn't happen over a night, a long time, a lot of time, a lot longer than the Grand Canyon, a very long time	Years, thousands of years, maybe millions, 100,000 to 1 million years, 50 to 100 million years

[1]Data are from 15 interviewees. One participant did not complete the GET. Another student's data were excluded from analysis due to explanations that continued to change even after follow-up probes. Columns two and three show range of qualitative and quantitative descriptions of durations but not frequencies. Terms like "hundreds of years" or "thousands of years" were placed in column 3, since these terms represent more specific temporal periods than "awhile" or "a long time," even though "hundreds of years" still represents a range of time periods.

mentioned in reference to the breakup of Pangea, the Grand Canyon, and coral reefs and was unanticipated, as in everyday contexts we often speak of elapsed time even while an event is still occurring (e.g., "I've been waiting half an hour for a table and don't know how much longer it will be till I'm seated.") It is difficult to know what to make of this, but it would be interesting to see if this idea is true of a broader spectrum of introductory geoscience students or if it is unique to this sample. A 21-y-old man talked about the Colorado River remaining in one place while tectonic plates were moving around it, perhaps indicating that he viewed only some portions of Earth as dynamic. The final notion was one Piaget (1969) found with young children, namely, the idea that if two events are still continuing, their durations must be similar, a view that considers only ending but not starting times.

> *And then the carving of the Grand Canyon. It's still being done, but so are the coral reefs, so I put them together really close. But they've been taking the longest because they're still changing shapes. (F8: 20-y-old female)*

Placement of 1 y relative to 100,000,000 y on task 4 was compared with placement of D (Earth's orbital period) relative to the event deemed to have the longest duration on the GET (task 5) to explore the relationship between students' ideas about the relative durations of temporal periods and their associated geoscience events. Eight students chose an event other than the breakup of Pangea as the one with the longest duration. The decision was made to use the event marked as having the longest duration for comparison, since that event represents those individuals' conceptions of the maximum duration of the items on the time line. Figure 1b shows task 4 for F6, the same student whose GET appears in Figure 1a. Even though she correctly stated Earth's rotational and orbital periods in her explanation for the placement of events, she spaced the events (D vs. C) farther apart on the GET than she did on task 4 (C vs. G). There is greater spacing between events up to 1 y on the GET than there is between corresponding time periods on task 4, and the three events with the longest durations on the GET were closer together than comparable time periods on task 4. Whether this is better explained by logarithmic mapping (Siegler and Opfer, 2003; Dehaene et al., 2008; Siegler et al., 2009) or as a proportional judgment task (Barth and Paladino, 2011) is unclear. She starred two events on the GET because she was unsure about their durations, which could explain why she placed them differently on the GET than she had on task 4. It does not explain why events with durations up to 1 y were spread farther apart on the GET.

To further investigate the comparison between task 4 and the GET, I measured the distances from the origin to 1 y and 100,000,000 y on task 4, and Earth's orbital period and the event with the longest duration on the GET. Ratios of the longer period to the smaller in each pair were calculated for all participants, and a mean ratio was determined. Ratios for "100,000,000 y: 1 y" (mean [M] = 8.56, SD = 5.43) and "event with longest duration: Earth's orbital period" (M = 5.39, SD = 4.93) coupled with the qualitative data in Table VI indicate there is little consensus among these students about the durations of the geologic processes considered here. There is also wide disagreement about the relationships among the numeric time periods. Mean ratios show that proportional relationships between temporal periods were severely underestimated in all cases. The imprecision of the task could be partly to blame, as mentioned earlier. Yet, it probably does not fully account for the results. The gross underestimation of proportional relationships between temporal periods underscores inferences made from the qualitative data, such as, as a group, the students in the sample (1) hold ideas about large numbers that are not sufficiently robust to deal with long temporal periods, and (2) find it difficult to relate the ideas they do have about numbers to specific geoscience markers.

GENERAL DISCUSSION

This paper attempted to show that at least some students in introductory university geoscience courses possess ideas about the durations of geologic processes that reflect more than just their GCK. These results show that students' tendency to equate spatial size with duration previously observed when considering relative age (Dodick and Orion, 2003, 2006) also applies to their thinking about temporal duration. The same propensity to use space as a marker for time when thinking about short and long time periods (Boroditsky, 2000; Boroditsky and Ramscar, 2002; Casasanto and Boroditsky, 2008; Casasanto et al., 2010; Cheek, 2011) suggests that the ways in which people reason about time are similar across temporal scales. The assumption that more space equals more time may be a basic one that is invoked in the absence of relevant knowledge that would preclude it, in this case GCK.

These students all knew the numbers on the time lines differed by orders of magnitude, but many could not relate the numbers to each other in anything more than a qualitative, intuitive way. When they used quantitative terms accurately, it was generally only for smaller time periods or those for which the multiplicative relationship was fairly transparent, such as between 1,000 and 100,000. Spatial compression of large time periods was common, similar to what others have found with cardinal numbers (Dehaene et al., 2008). Students were confused about the relative size of the units *thousands* and *millions*. If those units are viewed as merely very large time periods, it is questionable whether students will attach meaning to hearing that a process occurred over a period of 120 million years or how that compares to a process that occurs over 100,000 y or one that occurs over hundreds of years.

The primarily qualitative language used to describe the durations of geoscience events (GET) may imply that when students think about events such as the carving of the Grand Canyon by the Colorado River, the amount of time required for it to occur is not part of their thought processes. This could explain why Lee et al. (2011) found that college students were largely inaccurate when judging durations of science processes outside human timescales. Even individuals, like F6, who displayed a linear proportional (LP) understanding of large time periods found it difficult to associate those periods with specific geoscience events (see Fig. 1). F6 is a good example of someone who may benefit from linking events and processes to specific temporal duration markers, as has previously been suggested for temporal succession (Trend, 2001; Delgado, 2013). That recommendation would be less likely to help the 12 students classified as intuitive, nonlinear (IN) as they demonstrated confusion about the numbers themselves apart from their associated geoscience content.

It might be argued that the findings reported here are artifacts of the study's sample. Perhaps students from a religiously conservative institution have greater difficulty with long time periods since they are more likely to espouse young Earth creationism than students from secular institutions. A young Earth creationism position would limit students' view of the past, but it would not necessarily imply that they see time periods in the millions as impossible. The institution itself has no official statement on the age of Earth, but it does have an articulated position on the future (i.e., eternity), which is described as boundless, extending well beyond 4.5 or even 13.7 By. The decoupling of temporal periods from specific geoscience referents in several tasks provided opportunity to investigate whether ideas about large numbers operate independently of GCK. Young Earth creationism views would have been most likely to impact responses for task 5, the GET. There are hints of that in Table VI.

The meaning students attached to the rock-type patterns on the line drawing of sedimentary strata was unexpected. Block diagrams are included in all introductory geology texts, with generally no discussion about the patterns or their meaning. Even when rock types are identified, there is no attention to the misconceptions novices may be developing about the rocks themselves or about how they are formed based upon the patterns used in the diagrams. If these participants are typical, discussion of the meaning of the patterns in introductory geoscience courses is recommended. Curriculum developers might include explanatory information that would help novices interpret figures accurately.

STUDY LIMITATIONS

This study has a number of limitations. The same person conducted interviews and analyzed data, with the exception of the second coder, who independently sorted time lines. There is the possibility of inconsistency or bias in data analysis. Because time lines depended upon interviewees' drawing skills and verbal explanations, some people may have been misclassified on the basis of their drawing or verbal skills relative to their peers. Some people classified as IN may understand large numbers well, but neither their time lines nor their explanations are good measures of their true understanding. Furthermore, the criteria used to sort time lines constrain the conclusions reached. Participant motivation could have influenced responses on animations and time lines alike, as some may have decided answers were "good enough." Another possibility is that the tasks may have been so unusual that students would have benefited from opportunities to practice the task before being asked about their conceptions. Future studies should include a series of practice trials for the animations and a practice time line for more familiar temporal periods, such as those up to 100 y. Sample size and the biased nature of the population from which it was drawn could also have affected results, particularly for tasks that dealt with geoscience content. Results reported here need to be more widely replicated to see if they hold true for larger, more representative samples.

CONCLUSIONS AND RECOMMENDATIONS

It is unwise to assume all university undergraduates enter introductory geoscience courses with sufficient concepts of large numbers and their proportional relationships to understand long temporal durations. They may also not be applying ideas about the relationships among rate, size, and duration that they know to be true for shorter temporal scales to longer ones.

Linking temporal periods to specific geoscience processes is essential because the rates at which processes occur and their temporal durations are important components of GCK. However, it may not be enough. Addressing underlying ideas about time/space and numbers may be required in order for students to be able to place geoscience content in a temporal context. An assumption that all students are making those connections themselves is probably faulty. Explicit direct instruction about large numbers, time, and space is probably required.

A fuller exploration of the relationship between spatial reasoning and ideas about temporal durations of geologic processes could result in more effective use of spatial metaphors to improve temporal reasoning in geoscience. Instructors may need to point out to students everyday examples in which space and time do not covary and guide them to see how that is relevant for inferring durations of geologic processes. The importance of spatial reasoning to many other aspects of geoscience learning has been acknowledged for some time. Linking long temporal durations in relativistic ways to shorter time periods could provide scaffolding for students when learning about the duration of geologic processes. What seems obvious to us may not be to our students (Liben et al., 2011). In addition, discussions about the temporal durations of geoscience processes should accompany instruction about the processes. The fact that students transferred knowledge from the animations to the stratigraphic sequence and used more geoscience vocabulary after watching them is intriguing. Simple animations like these may prove to be a useful tool to help students think about the relationships among rate, size, and duration. It is an area requiring further exploration. Competing views regarding reasons for nonlinear mapping of numbers were mentioned in the Introduction, but the instrument used did not address whether difficulties comprehending long temporal periods are better explained by a logarithmic to linear mapping shift or are related to proportional reasoning skills. That is a question that can profitably be explored in future geoscience conceptions research.

Minimally, geoscience instructors could provide students with materials that visually show the relative relationships among large temporal periods, but, here again, explicit direct instruction about proportional relationships may be required. This is probably also true for instructors teaching about large spatial scale phenomena, such as in astronomy, where a poor understanding of large numbers could similarly impact learning. These results have implications for mathematics instruction at the K–12 level as well as within the university. In a world where large numbers are frequently in the news in many disciplines outside the geosciences (e.g., economics), number sense for very large (and very small) numbers is important. Helping students develop better number sense for large numbers may contribute to improved conceptual knowledge in a variety of domains.

Future research in this area is warranted. Research that investigates similar questions with other populations, including those from non-Western cultures, could provide data on the extent to which the observations made here are culturally dependent. Interventions that build upon the cognitive science literature described in the Introduction can explore how to use spatial metaphors to improve students' conceptions of long temporal durations. Simple animations like the ones used in this study have promise as possible teaching interventions for that purpose. Studies that investigate whether explicit teaching about large numbers and their proportional relationships has an effect on students' conceptions of long durations are needed.

Many students in introductory geoscience courses are there to fulfill general education requirements. Improving their understanding of the durations of geologic processes can contribute to a citizenry that will make informed choices for a sustainable future.

REFERENCES

Barth, H.C., and Paladino, A.M. 2011. The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, 14(1):125–135.

Booth, J., and Siegler, R. 2006. Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 41(6):189–201.

Boroditsky, L. 2000. Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, 75:1–28.

Boroditsky, L., and Ramscar, M. 2002. The roles of body and mind in abstract thought. *Psychological Science*, 13(2):185–189.

Casasanto, D., and Boroditsky, L. 2008. Time in the mind: Using space to think about time. *Cognition*, 106:579–593.

Casasanto, D., Fotakopoulou, O., and Boroditsky, L. 2010. Space and time in the child's mind: Evidence for a cross-dimensional asymmetry. *Cognitive Science*, 34:387–405, doi: 10.1111/j.1551–6709.2010.01094x.

Catley, K., and Novick, L. 2009. Digging deep: Exploring college students' knowledge of macroevolutionary time. *Journal of Research in Science Teaching*, 46(3):311–332.

Charmaz, K. 2006. Constructing grounded theory: A practical guide through qualitative analysis. Thousand Oaks, CA: Sage Publications.

Cheek, K.A. 2011. Exploring the relationship between students' understanding of conventional time and geologic time. *International Journal of Science Education*, 1–21, doi: 10.1080/09500693.2011.587032.

Cheek, K.A. 2012. Students' understanding of large numbers as a key factor in their understanding of geologic time. *International Journal of Science and Mathematics Education*, 10(5):1047–1069, doi: 10.1007/s10763-011-9312-1.

Confrey, J. 1991. Learning to listen: A student's understanding of powers of ten. *In* von Glaserfeld, E., ed., Radical constructivism in mathematics education. Dordrecht, The Netherlands: Kluwer Academic Publishers, p. 111–138.

Creswell, J.W. 2009. Research design: Qualitative, quantitative, and mixed methods approaches. Thousand Oaks, CA: Sage Publications.

Dahl, J., Anderson, S., and Libarkin, J. 2005. Digging into Earth science: Alternative conceptions held by K–12 teachers. *Journal of Geoscience Education*, 12:65–68.

Dehaene, S., Izard, V., Spelke, E., and Pica, P. 2008. Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, 320:1217–1220.

Delgado, C. 2013. Navigating deep time: Landmarks for time from the Big Bang to the Present. *Journal of Geoscience Education*, 61:103–112.

Dodick, J., and Orion, N. 2003. Cognitive factors affecting student understanding of geologic time. *Journal of Research in Science Teaching*, 40(4):415–442.

Dodick, J., and Orion, N. 2006. Building an understanding of geological time: A cognitive synthesis of the "macro" and "micro" scales of time. *In* Manduca, C., and Mogk, D., eds., Earth and mind: How geologists think and learn about the Earth. Special Paper 413. Boulder, CO: Geological Society of America, p. 77–94.

Friedman, W. 1992. Time memory and time perception. *In* Macar, F., Pouthas, V., and Friedman, W., eds., Time, action and cognition: Towards bridging the gap. Dordrecht, The Netherlands: Kluwer Academic Publishers, p. 165–172.

Gentner, D., Imai, M., and Boroditsky, L. 2002. As time goes by: Evidence for two systems in processing space-time metaphors. *Language and Cognitive Processes*, 17(5):537–565.

Hidalgo, A., and Otero, J. 2004. An analysis of the understanding of geological time by students at secondary and post-secondary level. *International Journal of Science Education*, 26(7):845–857.

Izard, V., and Dehaene, S. 2008. Calibrating the mental number line. *Cognition*, 106:1221–1247.

Kadosh, R., Tzelgov, J., and Henik, A. 2008. A synthetic walk on the mental number line: The size effect. *Cognition*, 106:548–557.

Kastens, K.A., Agrawal, S., and Liben, L.S. 2009. How students and field geologists reason in integrating spatial observations from outcrops to visualize a 3-D geological structure. *International Journal of Science Education*, 31(3):365–393.

Kortz, K.M., and Murray, D.P. 2009. Barriers to college students learning how rocks form. *Journal of Geoscience Education*, 57(4):300–315.

Lee, H.-S., Liu, O.L., Price, C.A., and Kendall, A.L.M. 2011. College students' temporal-magnitude recognition ability associated with durations of scientific changes. *Journal of Research in Science Teaching*, 48(3):317–335.

Libarkin, J., Kurdziel, J., and Anderson, S. 2007. College student conceptions of geological time and the disconnect between ordering and scale. *Journal of Geoscience Education*, 55(5):413–422.

Liben, L.S., Kastens, K.A., and Christensen, A.E. 2011. Spatial foundations of science education: The illustrative case of instruction in introductory geological concepts. *Cognition and Instruction*, 29(1):45–87, doi: 10.1080/07370008.2010.533596.

Marin, N. 2004. How can we identify replies that accurately reflect students' knowledge? A methodological proposal. *International Journal of Science Education*, 26(4):425–445.

McPhee, J. 1981. Annals of the former world. New York: Farrar, Straus, and Giroux.

Myers, M., and Newman, M. 2007. The qualitative interview in IS research: Examining the craft. *Information and Organization*, 17:2–26.

National Research Council. 2011. A framework for K–12 science education: Practices, crosscutting concepts, and core ideas. Washington, D.C.: National Academies Press.

Orion, N., and Ault, C., Jr. 2007. Learning Earth sciences. *In* Abell, S., and Lederman, N., eds., Handbook of research on science education. Mahwah, NJ: Lawrence Erlbaum Associates, p. 653–687.

Ozdemir, G. 2010. Exploring visuospatial thinking in learning about mineralogy: Spatial orientation and spatial visualization ability. *International Journal of Science and Mathematics Education*, 8(4):737–759.

Petitto, A. 1990. Development of numberline and measurement concepts. *Cognition and Instruction*, 7(1):55–78.

Piaget, J. 1969. The child's conception of time. New York: Ballantine Books.

Posner, G., and Gertzog, W. 1982. The clinical interview and the measurement of conceptual change. *Science Education*, 66(2):195–209.

Ritger, S.D., and Cummins, R.H. 1991. Using student-created metaphors to comprehend geologic time. *Journal of Geological Education*, 39:9–11.

Siegler, R., and Opfer, J. 2003. The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, 14(3):237–243.

Siegler, R., Thompson, C.A., and Opfer, J. 2009. The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education*, 3(3):143–150.

Stavy, R., and Tirosh, D. 2000. How students (mis-)understand science and mathematics intuitive rules. New York: Teachers College Press.

Swarat, S., Light, G., Park, E.J., and Drane, D. 2011. A typology of undergraduate students' conceptions of size and scale: Identifying and characterizing conceptual variation. *Journal of Research in Science Teaching*, 48(5):512–533, doi: 10.1002/tea.20403.

Teed, R., and Slattery, W. 2011. Changes in geologic time understanding in a class for preservice teachers. *Journal of Geoscience Education*, 59(3):151–162, doi: 10.5408/1.3604829.

Thompson, C.A., and Opfer, J. 2010. How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development*, 81(6):1768–1786.

Thompson, C.A., and Siegler, R. 2010. Linear-numerical-magnitude representations aid children's memory for numbers. *Psychological Science*, 21(9):1274–1281, doi: 10.1177/0956797610378309.

Trend, R. 1998. An investigation into understanding of geological time among 10- and 11-year-old children. *International Journal of Science Education*, 20(8):973–988.

Trend, R. 2000. Conceptions of geological time among primary teacher trainees, with reference to their engagement with geoscience, history, and science. *International Journal of Science Education*, 22(5):539–555.

Trend, R. 2001. Deep time framework: A preliminary study of U.K. primary teachers' conceptions of geological time and perceptions of geoscience. *Journal of Research in Science Teaching*, 38(2):191–221.

Walsh, V. 2003. A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7(11):483–488.

Yin, R.K. 2011. Qualitative research from start to finish. New York: The Guilford Press.

APPENDIX 1. Initial and category codes for reasons cited for answers on animations (task 2).

Sample Statements from Transcripts	Initial Code	Category Code
"I would say their size and because of that they took the same, just because of the size, how big they were exactly." "Because it's shorter and it took a shorter amount of time to fill."	Relative size of layers	A. Size directly proportional to duration
"Green layer went faster." "I was just watching and then it looked like the brown layer was slower than the green layer."	Rate/speed	B. Inverse proportionality of rate and duration
"I can see just by looking at it. I can't really explain how. I don't know." "I don't know for sure, but I feel like the red layer took a longer time than the blue."	Feeling/perception	C. Subjective, qualitative impression
"Just remembering which one took longer or shorter to fill."[1]	Memory	C. Subjective, qualitative impression
"Because as like the layers like filled I would write 0 to 5 and then I wrote an R next to it so I knew it was the red. And then I did that for every single one, so like blue was 5 to 13." "Cause I counted it."	Counting	D. Use of iterated, equivalent temporal units
"I timed it." "There's a stopwatch at the top."	Use of timer	D. Use of iterated, equivalent temporal units
"They're evenly proportional right now. In the other video they were all different like depths or whatever. The bottom one was actually a little bit faster than the 2nd layer." "They were filling at the same speed, but there's more blue than yellow."	Rate/speed and relative size of layers	E. Relationship of size and rate to duration are considered simultaneously when both vary

[1]Sample reasons in each row in the first column are from two different students, with the exception of the row initially coded "Memory." One student cited "memory" three times across all animations (twice for Animation 2 and once for Animation 3). Statements in column one are students' final statements regarding their reasons after any follow-up questions were asked.